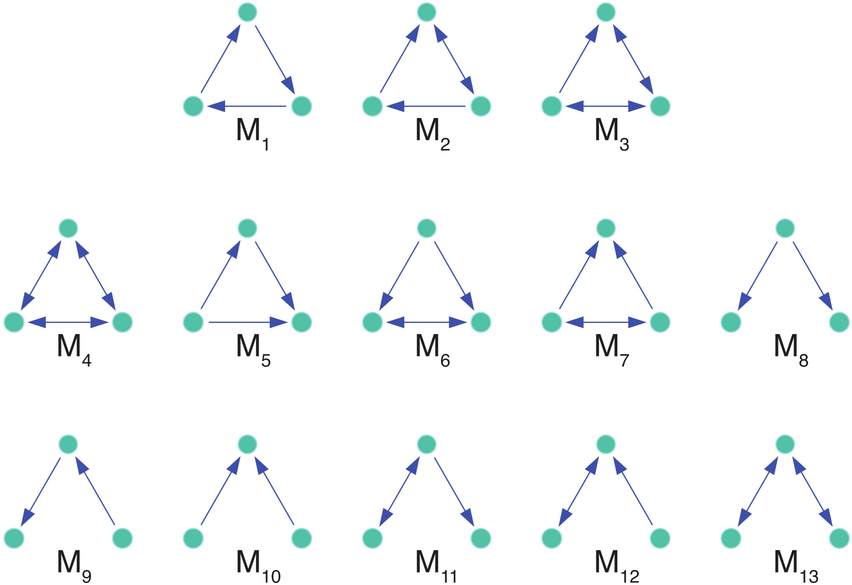
Higher-Order Organization of Complex Networks

Austin Benson, David Gleich,
Jure Leskovec

Higher-Order Structures

Small subgraphs (motifs, graphlets) are building blocks of networks



Higher-Order Structures

Higher-order structures are fundamental for control and function of complex systems:

- Social networks [Watts, Strogatz ‘98]
- Metabolic networks [Milo et al. ‘02]
- Protein networks [Alon ‘07] [Vinayagama et al.’16]
- Transportation networks [Rosvall ‘14]
- Neural networks [Park, Friston ‘13]
- Food webs [Stouffer et al. ‘12]

Higher-Order Organization

Prior work is based on simply counting frequencies of individual subgraphs

We still lack understanding of higher-order organization of complex systems

This talk: Higher-order Organization

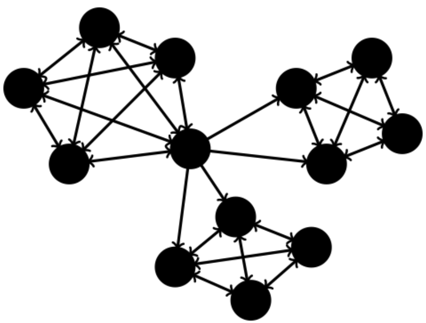
What is higher order organization of networks?

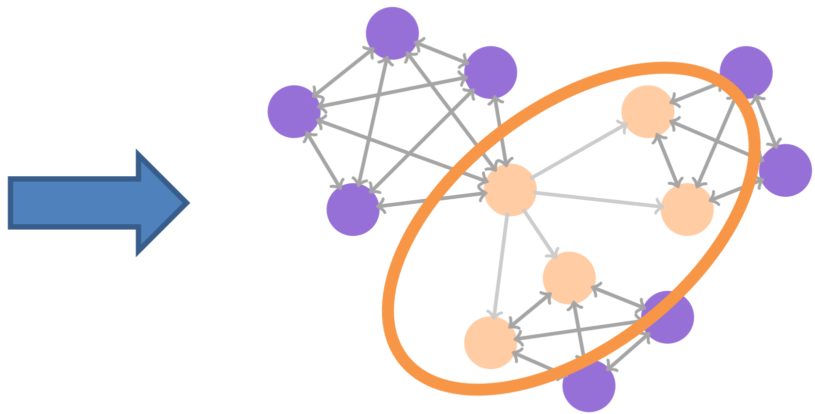
How do we extract it?

This talk: Modules of Motifs

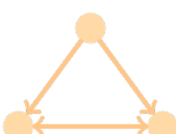
Find modules based on motifs!

Network:



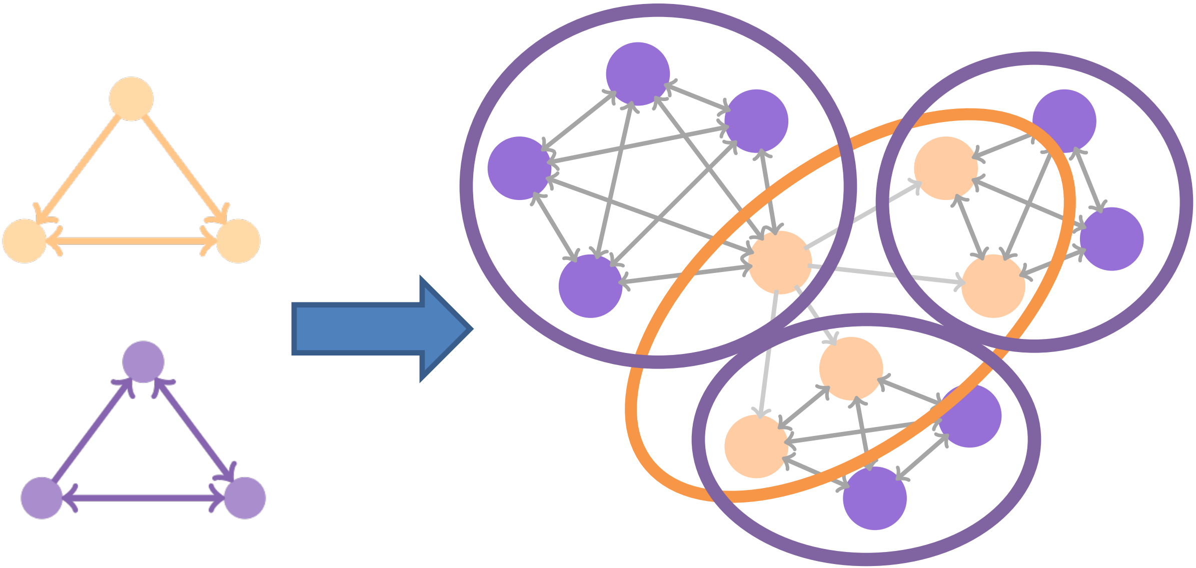


Motif:



[Serrour, Arenas, Gomez '11]

Modules of Motifs

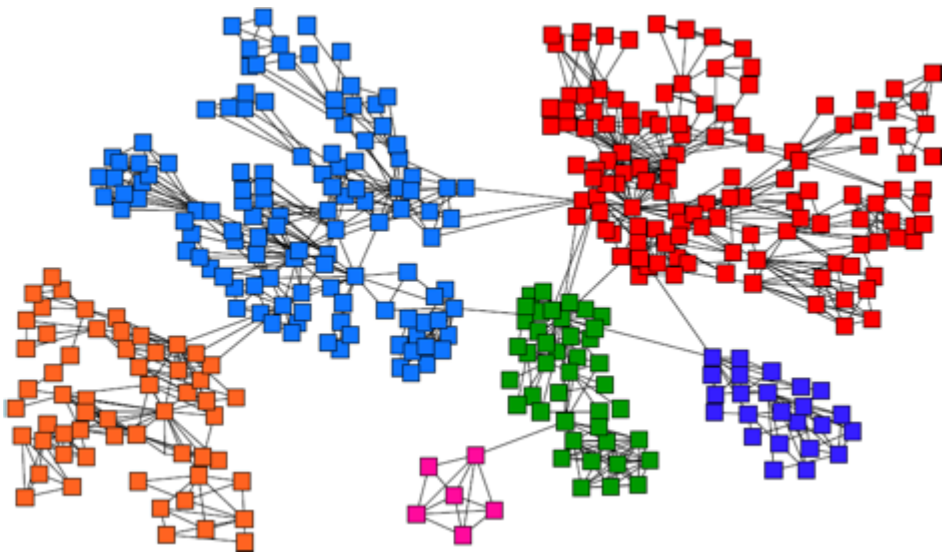


Different motifs reveal different modular structures!

How do we find modules of network motifs?

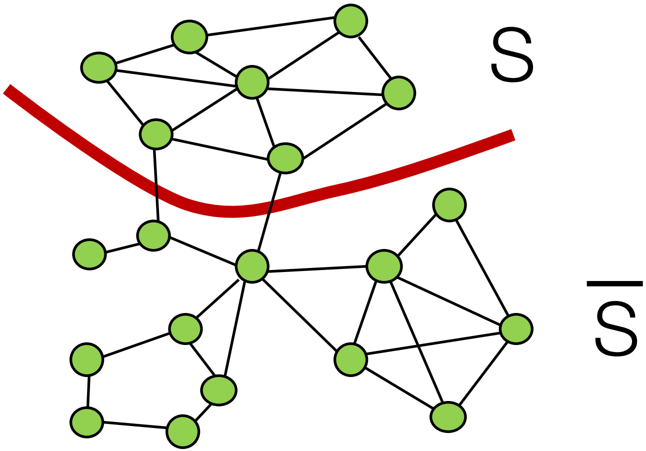
Background: Communities

- Old idea: Find densely connected sets of nodes



Background: Communities

- Define an objective function $\phi(S)$

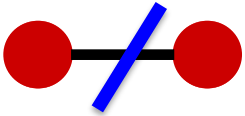


- Good community S cuts few edges while keeping many inside the set S

Defining an Objective Function

Conductance of a cluster S:

$$\phi(S) = \frac{\#(edges\ cut)}{vol(S)}$$

$edges\ cut$: 

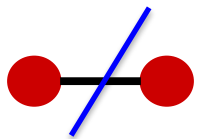
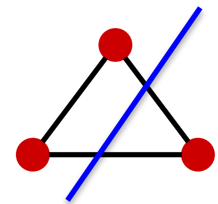
$vol(S)$ = #(edge end points in S)

Lower conductance means a better cluster

How do we generalize conductance to network motifs?

Defining Motif Conductance

Generalize Cut and Volume to motifs:

$edges\ cut$  → $motifs\ cut$ 

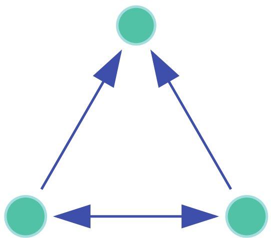
$vol(S)$ = #(edge end-points in S) → $vol_M(S)$ = #(motif end-points in S)

$$\phi(S) = \frac{\#(edges\ cut)}{vol(S)} \rightarrow \phi(S) = \frac{\#(motifs\ cut)}{vol_M(S)}$$

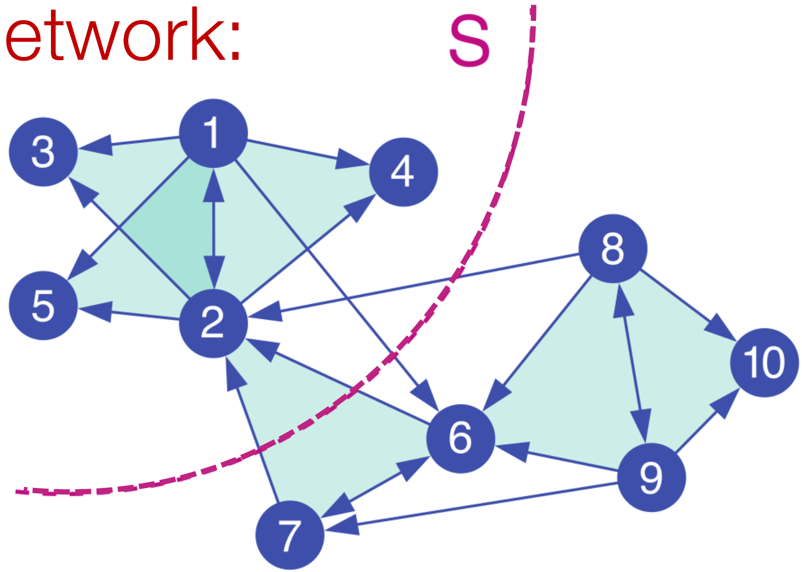
[Benson, Gleich, Leskovec, Science 2016]

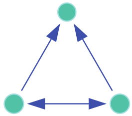
Motif Conductance: Example

Motif:



Network:





$$\phi_M(S) = \frac{\text{motifs cut}}{\text{motif volume}}$$

Higher-order Partitioning

How do we find clusters of motifs?

- Given a motif M and a graph G
- Find a set of nodes S that minimizes motif conductance

Bad news: Finding set S with the minimal motif conductance is computationally intractable (NP-hard)!

[Benson, Gleich, Leskovec, Science 2016]

Motif Spectral Clustering

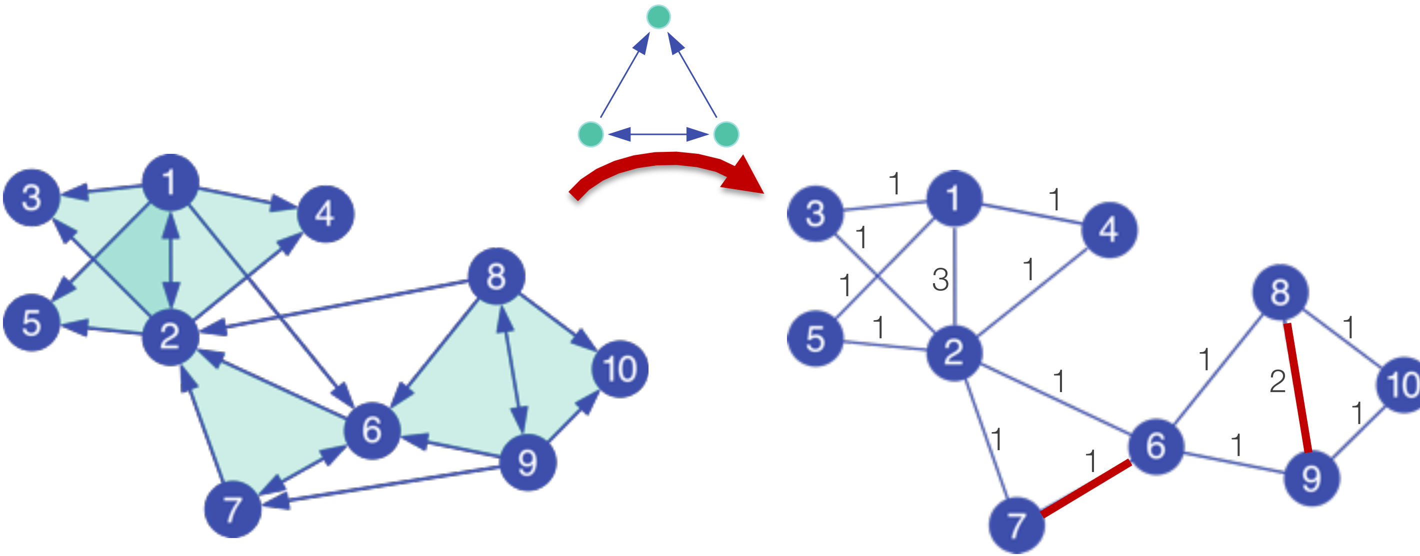
Solution: Motif Spectral Clustering

- Input: Graph G and motif M
- Using G form a new weighted graph W
- Apply spectral clustering on W
- Output the clusters

Theorem: Resulting clusters will obtain near optimal motif conductance

[Benson, Gleich, Leskovec, Science 2016]

Motif Spectral Clustering



Graph G

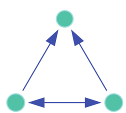
Weighted graph $W^{(M)}$

$W_{ij}^{(M)}$ = # of times edge (i,j) participates in motif M

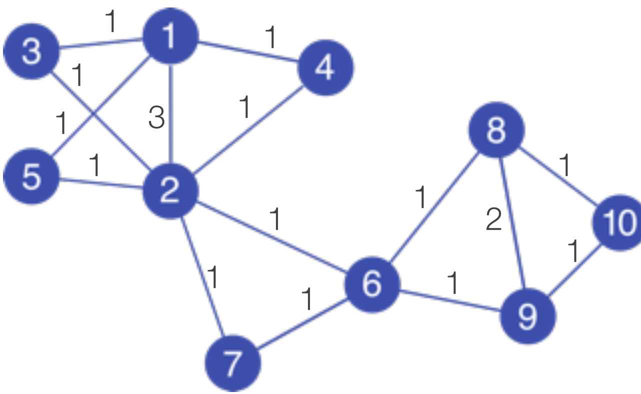
Motif Spectral Clustering

Insight:

Spectral clustering on weighted graph $W^{(M)}$ finds clusters of low motif conductance:



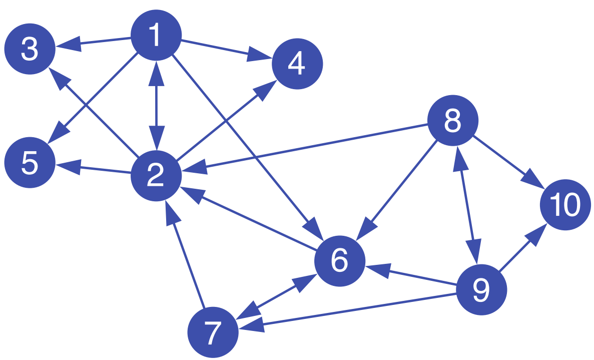
$$\phi_M(S) = \frac{\text{motifs cut}}{\text{motif volume}}$$



Weighted graph $W^{(M)}$

$W_{ij}^{(M)}$ = # of times edge (i,j) participates in motif M

Motif Spectral Clustering: 1



Nodes (columns) × Nodes (rows)

	1	2	3	4	5	6	7	8	9	10
1	0	**3**	1	1	1	0	0	0	0	0
2	**3**	0	1	1	1	1	1	0	0	0
3	1	1	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	1	1	1	0
7	0	1	0	0	1	0	0	0	0	0
8	0	0	0	0	0	1	0	0	**2**	1
9	0	0	0	0	0	1	0	**2**	0	1
10	0	0	0	0	0	0	0	1	1	0

Step 1: Form weighted graph $W^{(M)}$

Motif Spectral Clustering: 2

$$\mathcal{L}^{(M)} = D^{-1/2}(D - W^{(M)})D^{-1/2}$$

$$\mathcal{L}^{(M)}\mathbf{z} = \lambda_2\mathbf{z}$$

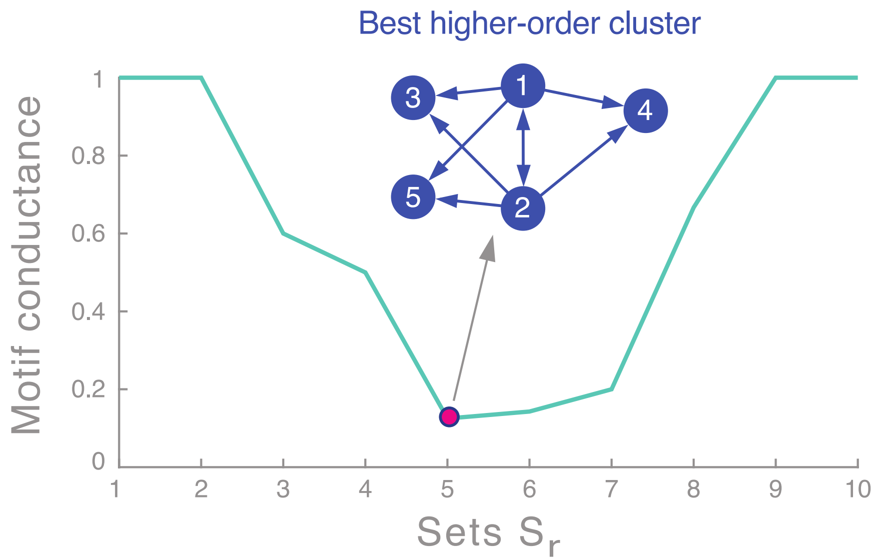
$$\mathbf{f}^{(M)} = D^{-1/2}\mathbf{z}$$

$D = \text{diag}(W^{(M)}\mathbf{e})$

Diagonal degree matrix

Step 2: Compute Fiedler vector $f^{(M)}$ associated with λ_2 of the Laplacian of $W^{(M)}$

Motif Spectral Clustering: 3



Step 3: Sort nodes by values in $f^{(M)}$: $f_1, f_2, \ldots f_n$

Let $S_r = \{f_1, \ldots, f_r\}$ and compute the motif conductance of each S_r

Motif Cheeger Inequality

Theorem: The algorithm finds a set of nodes S for which

$$\phi_M(S) \leq 4\sqrt{\phi_M^*}$$

$\phi_M(S)$… motif conductance of S found by our algorithm

ϕ_M^* … motif conductance of optimal set S*

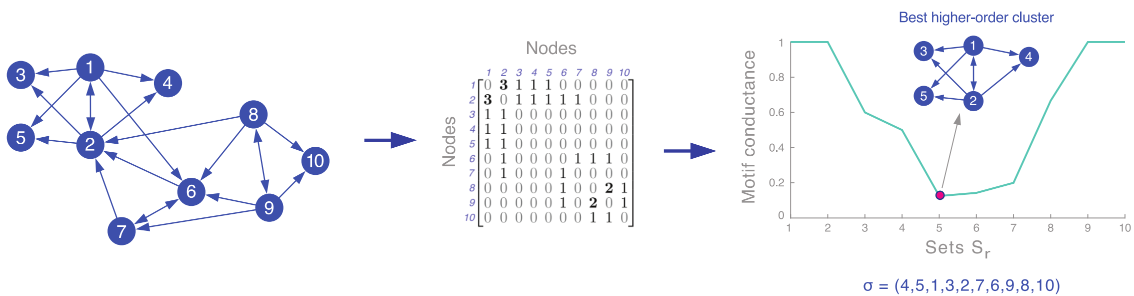
In other words: Clusters S found by our method are provably near optimal

Advantages

- Easy to implement: ~10 lines of Matlab
- Scalable to large networks (1.4B edge graphs)
- General: Can be applied to directed, signed, and weighted networks
- Code available: http://snap.stanford.edu/higher-order/

Summary

- Generalization of community detection to higher-order structures
- Motif-conductance objective admits a motif Cheeger inequality
- Simple, fast, and scalable:



Applications

1) We don’t know a motif of interest
   - Food webs and new applications

2) We know the motif of interest
   - Regulatory transcription networks, connectome, social networks

3) We seek richer information from data
   - Transportation networks

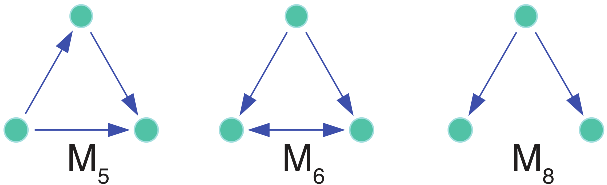
Application 1:
We do not know
the motif of interest

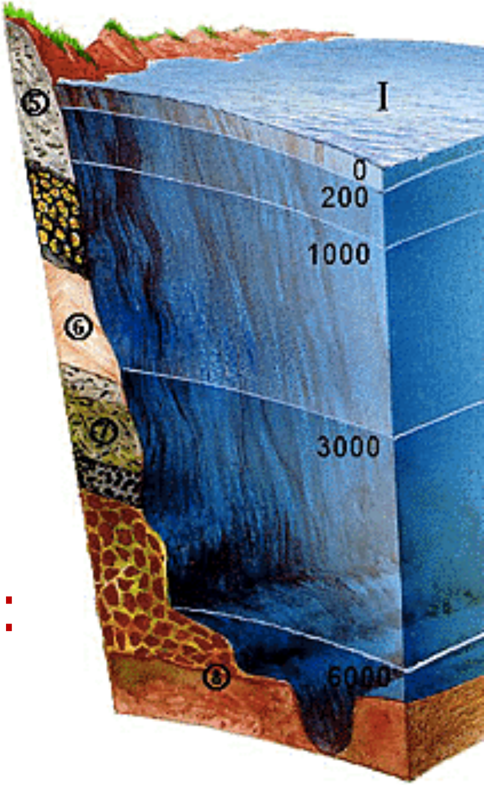
Application 1: Food webs

Florida Bay food web:

- Nodes: species in the ecosystem
- Edges: carbon exchange (who eats whom)

Different motifs capture different energy flow patterns:





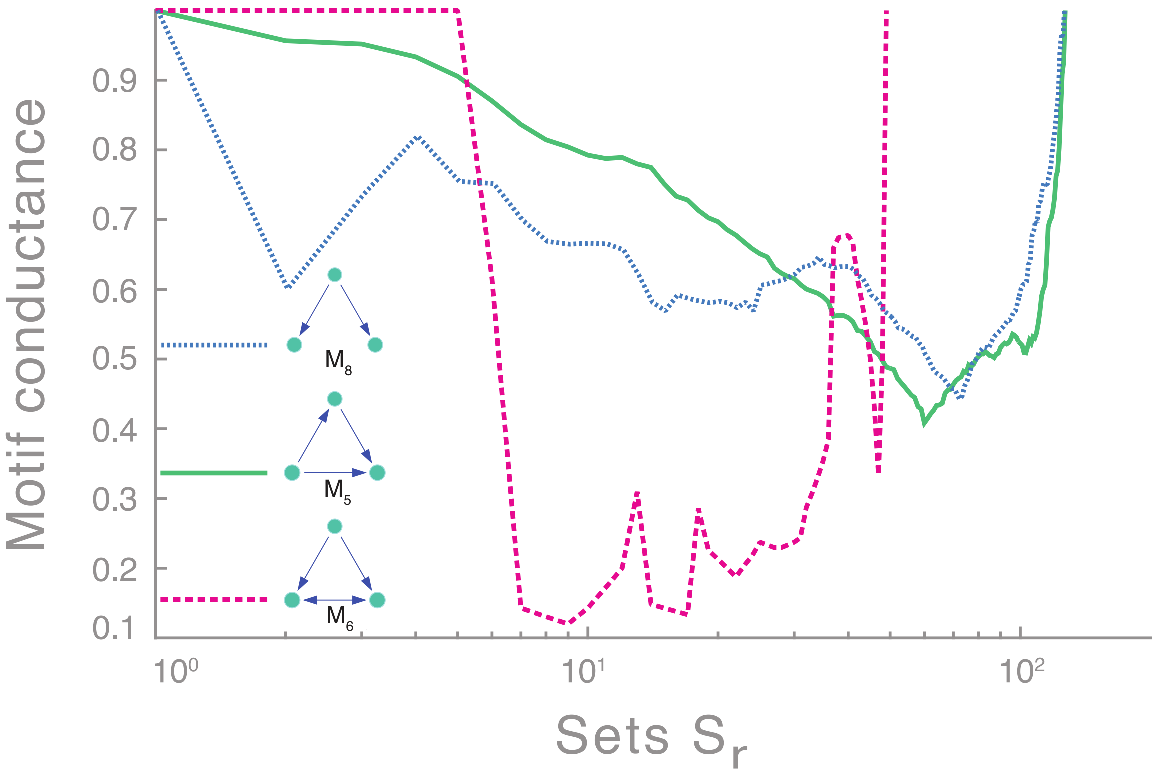
Florida Bay Food Web

Which motif organizes the food web?

Approach:

- Run motif spectral clustering separately for each of the 13 motifs
- Examine the Sweep profile to see which motif gives best clusters

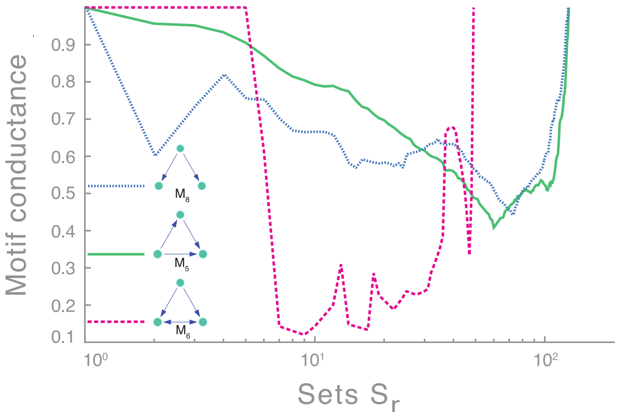
Florida Bay Food Web

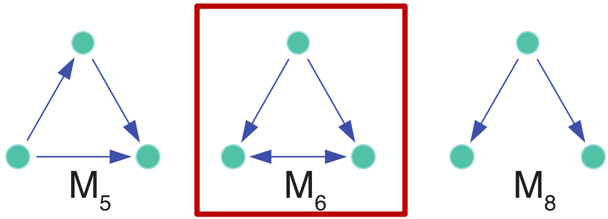


Food Web: Observations

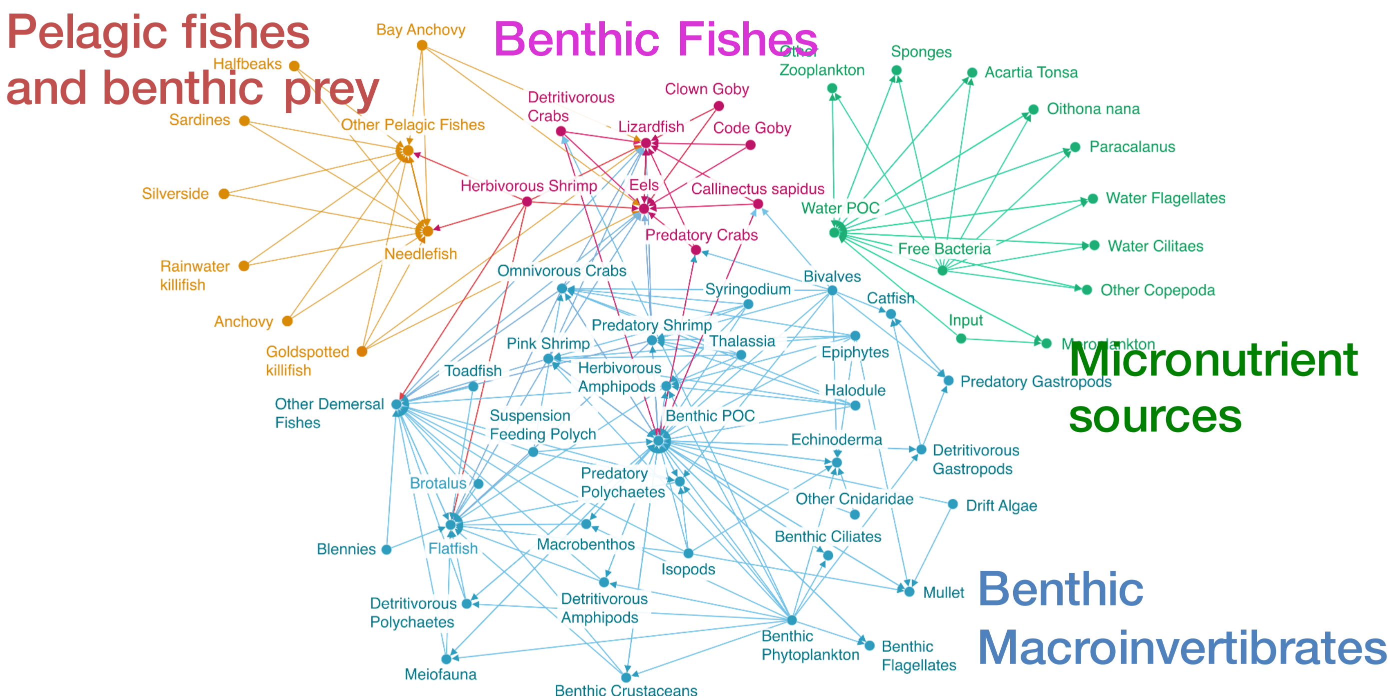
Observation:
Network organizes based on motif M_6 (but not M_5 or M_8)

- There exist good cuts for M_6 but not for M_5 or M_8



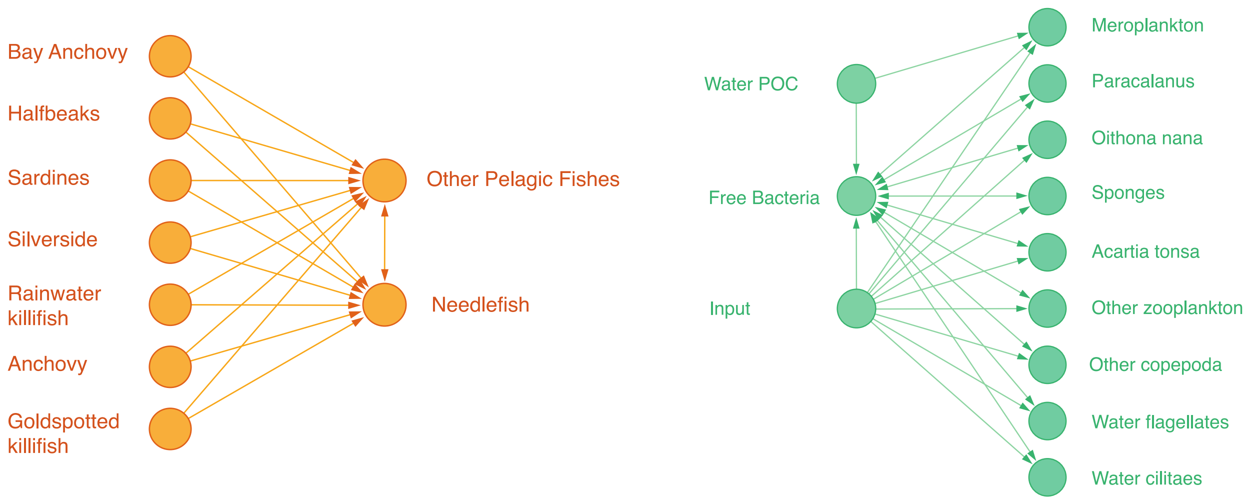


Food Web: Clusters



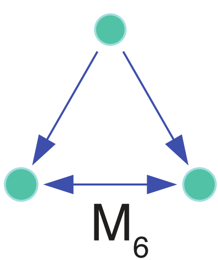
M_6 reveals known aquatic layers with higher accuracy (84% vs. 65%)

Structure of Aquatic Layers



Aquatic layers organize based on M_6

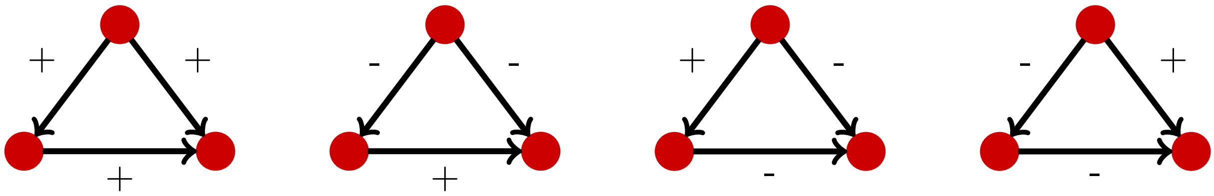
- Many instances of M_6 inside
- Few instances of M_6 cross



Application 2:
Signed networks

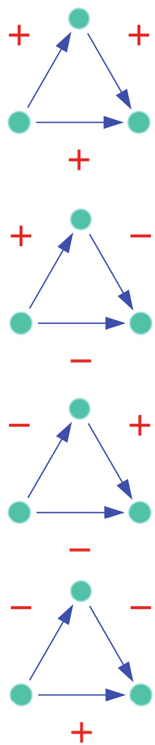
(2) Gene Regulatory Networks

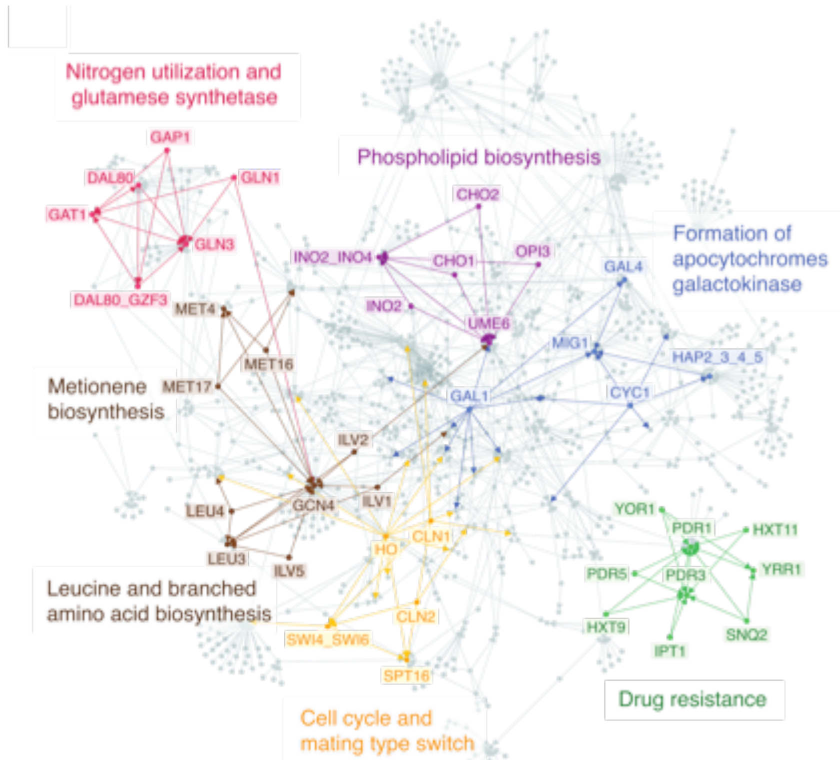
- Nodes are groups of genes in mRNA
- Edges are directed transcriptional regulation relationships



- The "feedforward loop" motif represents biological function [Alon '07]

Yeast Regulatory Network

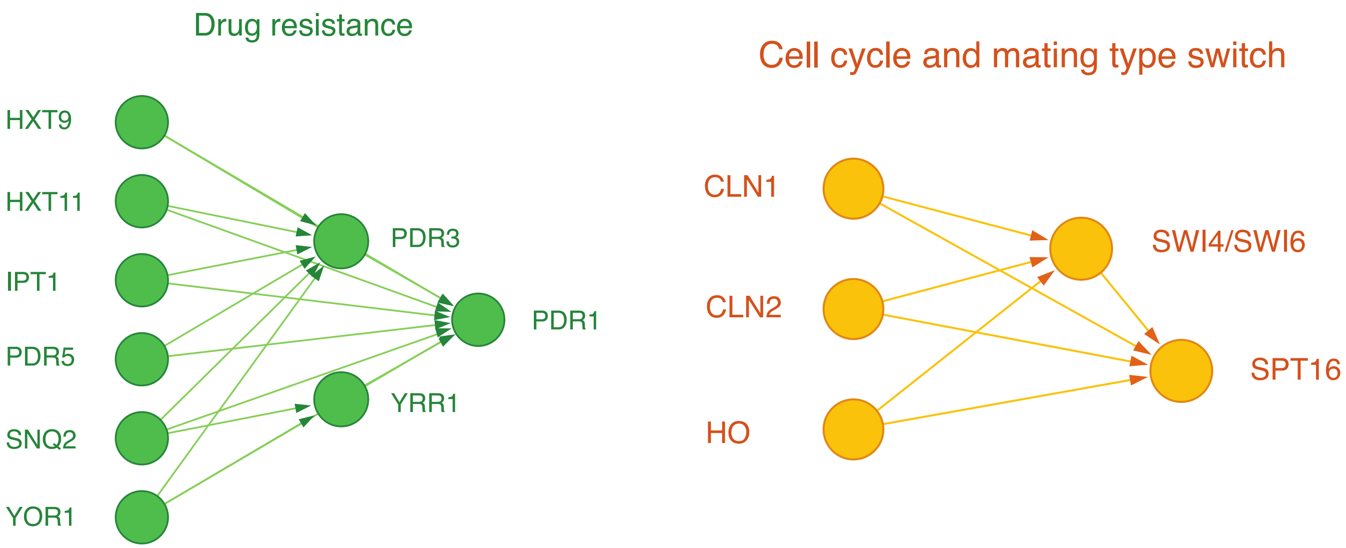




97% detection accuracy (vs. 68-82%)

Structure of Modules

- Feed forward loops:

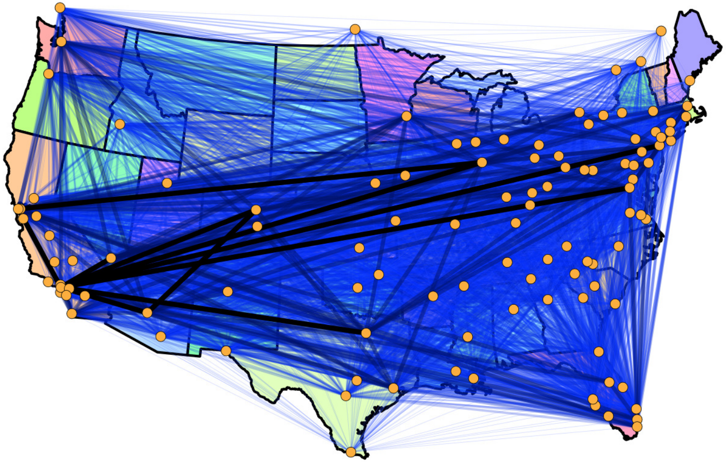


Application 3:
Exploring richer information from data

(4) Transportation Networks

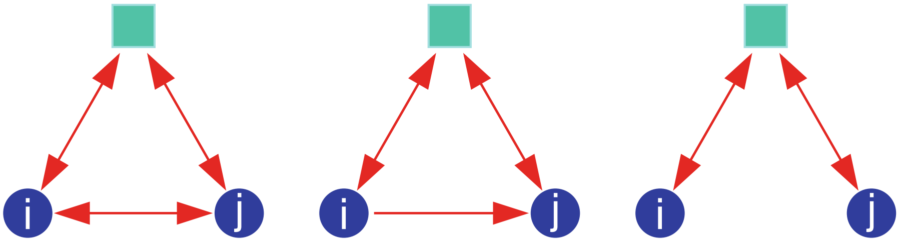
North American air transport network:

- Nodes are airports
- Unweighted directed edges reflect reachability
- (Based on Freyet al.'s 2007)



Transportation Networks

- Want to understand hub-and-spoke structure of the network
- Use "anchored" motifs:

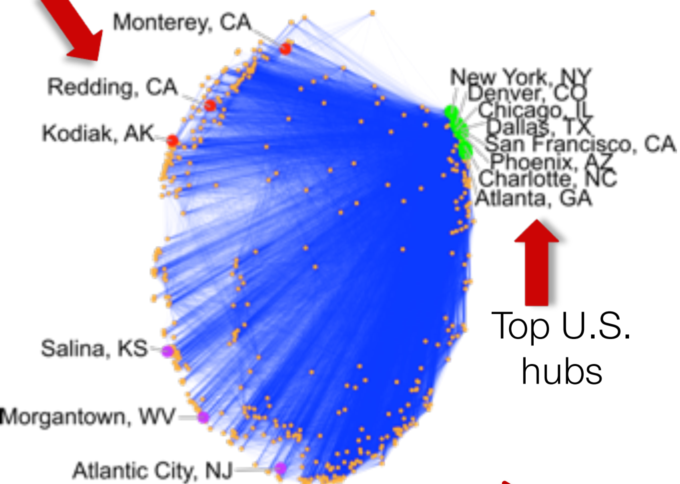


$(W_M)_{ij}$ = #{bidirectional 2-paths from i to j}

Transportation Networks

Motif-based spectral embedding

West coast non-hubs



Top U.S. hubs

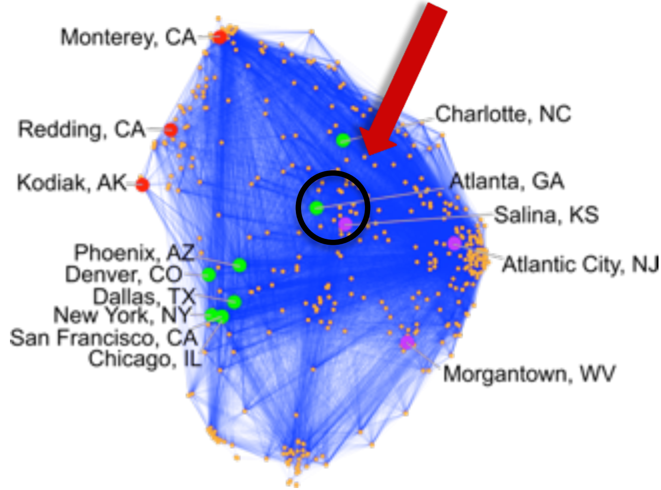
Secondary spectral coordinate

Primary spectral coordinate

East coast non-hubs

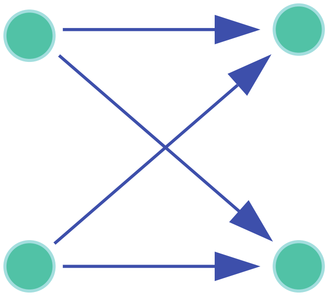
Edge-based spectral embedding

Atlanta, the top hub, is next to Salina, a non-hub.



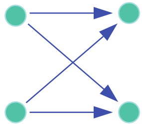
(3) Neuronal Networks

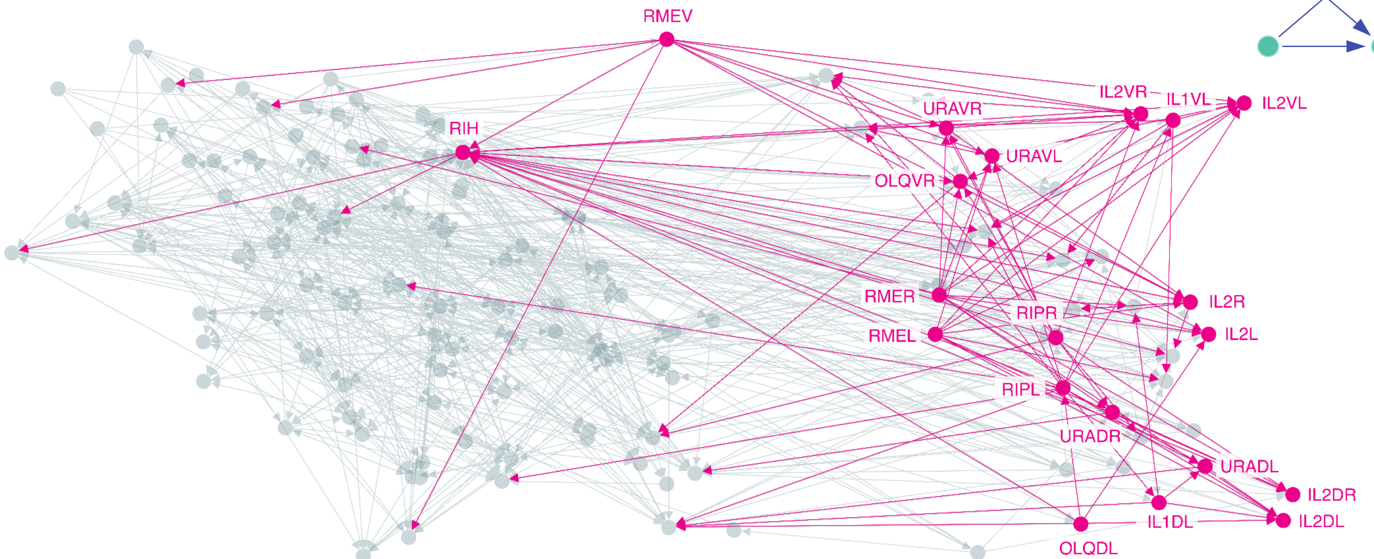
Bi-Fan motif is overexpressed in *C. Elegans* network [Milo '02]



Cluster the network based on this motif!

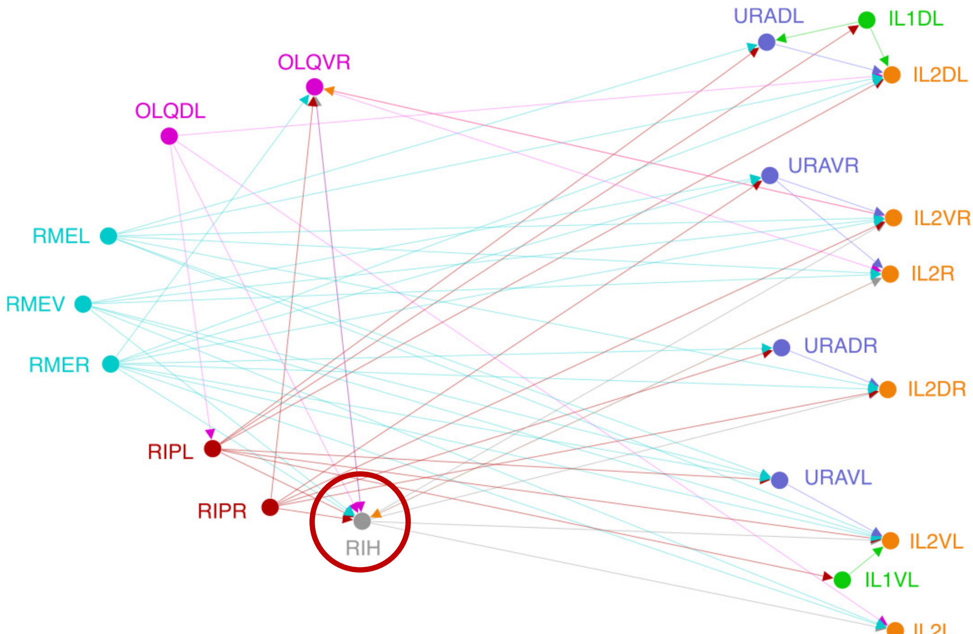
C. Elegans: Best Cluster





Node locations represent spatial coordinates of corresponding neurons

C. Elegans Network



- Ring motor (RME*) neurons act as inputs
- Inner labial sensory (IL2*) neurons are the destinations
- URA neurons act as intermediaries

Summary

- Community detection of higher-order structures
- Provably finds good clusters
- Simple, fast, and scalable
- Can be applied to directed, signed, and weighted networks

Paper, Data, Code

- Paper:
  Higher-order Organization of Complex Networks. Austin R. Benson, David F. Gleich, and Jure Leskovec. Science, 2016.

- Code & Data:
  http://snap.stanford.edu/higher-order/